

06323435 89339704 PMID: 2668181 Record Identifier: 89339704

Type A
Immunological characterization of papain-induced fragments of Clostridium botulinum type A neurotoxin and interaction of the fragments with brain synaptosomes.

Kozaki S ; Miki A; Kamata Y; Ogasawara J; Sakaguchi G

Department of Veterinary Science, College of Agriculture, University of Osaka Prefecture, Japan.

Infection and immunity (UNITED STATES) Sep 1989 , 57 (9) p2634-9,

ISSN 0019-9567 Journal Code: 0246127

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Other Citation Owner: NASA

Record type: Completed

Subfile: INDEX MEDICUS

After treatment of **Clostridium botulinum** type A neurotoxin with papain, three fragments (Mrs, 101,000, 45,000, and 43,000) were purified by hydrophobic and ion-exchange chromatography with a high-performance liquid chromatographic system. Immunoblotting analyses with monoclonal antibodies showed that the 101,000-dalton fragment consisted of the light chain and a part of the heavy chain (H-1 fragment) linked together by a disulfide bond, and the other two fragments were correlated to the remaining portion of the heavy chain (H-2 fragment). The 45,000- and 43,000-dalton fragments effectively competed for binding of the ¹²⁵I-labeled neurotoxin to synaptosomes, while no inhibition was observed with the 101,000-dalton fragment. The results indicate that the H-2 fragment interacts with the binding site on the neural membrane. The binding of the neurotoxin was impaired by treatment of synaptosomes with neuraminidase. Incorporation of gangliosides into neuraminidase-treated synaptosomes resulted in the restoration of binding. The results suggest that gangliosides are one of the components of the toxin-binding site.

Tags: Animal; Support, Non-U.S. Gov't

Descriptors: **Botulinum Toxins--immunology--IM;** *Brain--metabolism--ME; ***Clostridium botulinum --immunology--IM;** *Neurotoxins--immunology--IM; *Papain--pharmacology--PD; *Synaptosomes--metabolism--ME; Antibodies, Monoclonal; Antigens, Bacterial--immunology--IM; Binding Sites, Antibody; Binding, Competitive; **Botulinum Toxins--metabolism--ME;** **Botulinum Toxins--pharmacology--PD;** Immunoblotting; Mice; Mice, Inbred BALB C; Molecular Weight; Neurotoxins--metabolism--ME; Neurotoxins--pharmacology--PD

CAS Registry No.: 0 (Antibodies, Monoclonal); 0 (Antigens, Bacterial); 0 (Binding Sites, Antibody); 0 (Botulinum Toxins); 0 (Neurotoxins)

Enzyme No.: EC 3.4.22.2 (Papain)

Record Date Created: 19890915

Record Date Completed: 19890915

*Updated
5/23/03
V66*

06610371 90235864 PMID: 2185020

The complete amino acid sequence of the Clostridium botulinum type A neurotoxin, deduced by nucleotide sequence analysis of the encoding gene.

Thompson D E ; Brehm J K; Oultram J D; Swinfield T J; Shone C C;
Atkinson T; Melling J; Minton N P

Division of Biotechnology, Centre for Applied Microbiology and Research,
Porton Down, England.

SEQ
Type A

European journal of biochemistry / FEBS (GERMANY, WEST) Apr 20 1990 ,
189 (1) p73-81, ISSN 0014-2956 Journal Code: 0107600

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Record type: Completed

Subfile: INDEX MEDICUS

A 26-mer oligonucleotide probe was synthesized (based on the determined amino acid sequence of the N-terminus of the Clostridium **botulinum** type A neurotoxin, BoNT/A) and used in Southern blot analysis to construct a restriction map of the region of the clostridial genome encompassing BoNT/A. The detailed information obtained enabled the cloning of the structural gene as three distinct fragments, none of which were capable of directing the expression of a toxic molecule. The central portion was cloned as a 2-kb Pvull-TaqI fragment and the remaining regions of the light chain and heavy chain as a 2.4-kb ScaI-TaqI fragment and a 3.4-kb HpaI-Pvull fragment, respectively. The nucleotide sequence of all three fragments was determined and an open reading frame identified, composed of 1296 codons corresponding to a polypeptide of 149 502 Da. The deduced amino acid sequence exhibited 33% similarity to tetanus toxin, with the most highly conserved regions occurring between the N-termini of the respective heavy chains. Conservation of Cys residues flanking the position at which the toxins are cleaved to yield the heavy chain and light chain allowed the tentative identification of those residues which probably form the disulphide bridges linking the two toxin subfragments.

Tags: Comparative Study

Descriptors: **Botulinum Toxins--genetics--GE;** ***Clostridium botulinum --genetics--GE;** ***Genes, Structural, Bacterial;** ***Neurotoxins--genetics--GE;** Amino Acid Sequence; Base Sequence; Cloning, Molecular; Escherichia coli --genetics--GE; Molecular Sequence Data; Oligonucleotide Probes --biosynthesis--BI; Sequence Homology, Nucleic Acid

Molecular Sequence Databank No.: GENBANK/X52066; GENBANK/UNKNOWN

CAS Registry No.: 0 (Botulinum Toxins); 0 (Neurotoxins); 0 (Oligonucleotide Probes)

Record Date Created: 19900606

Record Date Completed: 19900606

06880738 91120847 PMID: 2126206

Botulinum neurotoxin type A: sequence of amino acids at the N-terminus and around the nicking site.

DasGupta B R ; Dekleva M L

Food Research Institute, University of Wisconsin, Madison 53706.

Biochimie (FRANCE) Sep 1990 , 72 (9) p661-4, ISSN 0300-9084

Journal Code: 1264604

Contract/Grant No.: NS17742; NS; NINDS; NS24545; NS; NINDS

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Record type: Completed

Subfile: INDEX MEDICUS

Clostridium botulinum synthesizes the type A botulinum neurotoxin (NT) as a approximately 150 kDa single chain protein. Post-translational proteolytic processing yields a approximately 150 kDa dichain protein composed of a approximately 50 kDa light and approximately 100 kDa heavy chain , which has higher toxicity. Trypsin's action mimics the endogenous proteolytic processing. The proteolytic cleavages could occur at 4 sites. We have examined 2 such sites and defined the peptide sequences before and after proteolytic processing. The N-terminal residues of the newly synthesized approximately 150 kDa single chain NT, Pro-Phe-Val-Asn-Lys-, remain intact at the N-terminus of the approximately 50 kDa light chain generated either in the clostridial culture or in vitro with trypsin or with a protease purified from the homologous bacterial culture. The clostridial protease cleaves the single chain NT in vitro, at 1/3 the distance from its N-terminus, on the amino side of Gly of the sequence -Gly-Tyr-Asn-Lys-Ala-Leu-Asn-Asp-Leu- before cleaving the bond Lys-Ala at a slower rate. The data indicate that the dichain NT is formed in the bacterial culture in at least 2 steps. Cleavage at X-Gly produces a approximately 100 kDa heavy chain -like fragment which is then truncated; cleavage 4 residues downstream at Lys-Ala, and excision of the tetrapeptide Gly-Tyr-Asn-Lys, generates the mature heavy chain with Ala as its N-terminal residue. The approximately 100 kDa heavy chain generated in vitro, by nicking the single chain NT with trypsin, also has Ala-Leu-Asn- as the N-terminal residues.

Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Descriptors: *Amino Acid Sequence; *Botulinum Toxins--chemistry--CH; *Clostridium botulinum; Botulinum Toxins--genetics--GE; Clostridium botulinum--genetics--GE; Molecular Sequence Data; Protein Processing, Post-Translational

CAS Registry No.: 0 (Botulinum Toxins)

Record Date Created: 19910311

Record Date Completed: 19910311

08214580 94280521 PMID: 8011071

Covalent structure of botulinum neurotoxin type A: location of sulfhydryl groups, and disulfide bridges and identification of C-termini of light and heavy chains.

Kriegstein K G ; DasGupta B R; Henschen A H

Department of Molecular Biology and Biochemistry, University of California, Irvine 92717.

Journal of protein chemistry (UNITED STATES) Jan 1994 , 13 (1)

p49-57, ISSN 0277-8033 Journal Code: 8217321

Contract/Grant No.: NS17742; NS; NINDS

Document type: Journal Article

Languages: ENGLISH

Main Citation Owner: NLM

Record type: Completed

Subfile: INDEX MEDICUS

Botulinum neurotoxin Type A is synthesized by Clostridium botulinum as a approximately 150 kD single chain polypeptide. The posttranslational processing of the 1296 amino acid residue long gene product involves removal of the initiating methionine, formation of disulfide bridges, and limited proteolysis (nicking) by the bacterial protease(s). The mature dichain neurotoxin is made of a approximately 50-kD light chain and a approximately 100-kD heavy chain connected by a disulfide bridge. DNA derived amino acid sequence predicted a total of 9 Cys residues (Binz et al., 1990, J. Biol. Chem. 265, 9153-9158; Thompson et al., 1990, Eur. J. Biochem. 189, 73-81). Treatment of the dichain neurotoxin, dissolved in 6 M guanidine. HCl, with 4-vinylpyridine converted 5 Cys residues into S-pyridylethyl cysteine residues; but alkylation after mercaptolysis converted all 9 Cys residues in the S-pyridylethylated form. After confirming the predicted number of Cys residues by amino acid analysis, the positions of the 5 Cys residues carrying sulfhydryl groups and the 4 involved in disulfide bridges were determined by comparing the elution patterns in reversed-phase HPLC of the cyanogen bromide mixtures of the exclusively alkylated and the mercaptolyzed-alkylated neurotoxin. The chromatographically isolated components were identified by N-terminal amino acid sequence analysis. The HPLC patterns showed characteristic differences. The Cys residues predicted in positions 133, 164, 790, 966, and 1059 were found in the sulfhydryl form; Cys 429 and 453 were found disulfide-bridge connecting the light and heavy chains, and Cys 1234 and 1279 were found in an intrachain disulfide-bridge near the C-terminus in the heavy chain. (ABSTRACT TRUNCATED AT 250 WORDS)

Tags: Support, Non-U.S. Gov't; Support, U.S. Gov't, P.H.S.

Descriptors: *Botulinum Toxins--chemistry--CH; Amino Acid Sequence; Botulinum Toxins--biosynthesis--BI; Chromatography, High Pressure Liquid; Clostridium botulinum--metabolism--ME; Cyanogen Bromide; Disulfides --analysis--AN; Macromolecular Systems; Molecular Sequence Data; Neurotoxins--chemistry--CH; Peptide Fragments--chemistry--CH; Peptide Fragments--isolation and purification--IP; Sulfhydryl Compounds--analysis --AN; Trypsin

CAS Registry No.: 0 (Botulinum Toxins); 0 (Disulfides); 0 (Macromolecular Systems); 0 (Neurotoxins); 0 (Peptide Fragments); 0 (Sulfhydryl Compounds); 506-68-3 (Cyanogen Bromide)

Enzyme No.: EC 3.4.21.4 (Trypsin)

Record Date Created: 19940726

Record Date Completed: 19940726

ExPASy Home page	Site Map	Search ExPASy	Contact us	Swiss-Prot			
Hosted by NCSC US	Mirror sites:	Canada	China	Korea	Switzerland	Taiwan	
Search	Swiss-Prot/TrEMBL	<input type="button" value="▼"/>	for	<input type="text"/>		<input type="button" value="Go"/>	<input type="button" value="Clear"/>

NiceProt View of Swiss-Prot: P10844

[Printer-friendly view](#)[Quick BlastP search](#)

[\[General\]](#) [\[Name and origin\]](#) [\[References\]](#) [\[Comments\]](#) [\[Cross-references\]](#)
[\[Keywords\]](#) [\[Features\]](#) [\[Sequence\]](#) [\[Tools\]](#)

Note: most headings are clickable, even if they don't appear as links. They link to the user manual or other documents.

General information about the entry

Entry name	BXB_CLOBO
Primary accession number	P10844
Secondary accession number	P10843
Entered in Swiss-Prot in	Release 11, July 1989
Sequence was last modified in	Release 26, July 1993
Annotations were last modified in	Release 42, September 2003

Name and origin of the protein

Protein name	Botulinum neurotoxin type B [Precursor]
Synonyms	EC <u>3.4.24.69</u> BoNT/B Bontoxilysin B
Gene name	BOTB
From	<u>Clostridium botulinum</u> [TaxID: <u>1491</u>]
Taxonomy	Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae; Clostridium.

References

[1] SEQUENCE FROM NUCLEIC ACID. MEDLINE=92384550; PubMed=1514783; [<u>NCBI</u> , <u>ExPASy</u> , <u>EBI</u> , <u>Israel</u> , <u>Japan</u>] <u>Whelan S.M.</u> , <u>Elmore M.J.</u> , <u>Bodsworth N.J.</u> , <u>Brehm J.K.</u> , <u>Atkinson T.</u> , <u>Minton N.P.</u> : "Molecular cloning of the <i>Clostridium botulinum</i> structural gene encoding the type B neurotoxin and determination of its entire nucleotide sequence."; <i>Appl. Environ. Microbiol.</i> 58:2345-2354(1992).
[2] SEQUENCE OF <u>35-245</u> FROM NUCLEIC ACID. STRAIN=NCTC 7273; <u>Szabo E.A.</u> , <u>Pemberton J.M.</u> , <u>Desmarchelier P.M.</u> : Submitted (APR-1992) to the EMBL/GenBank/DDBJ databases.
[3] SEQUENCE OF <u>633-993</u> FROM NUCLEIC ACID. STRAIN=NCTC 7273; MEDLINE=94013372; PubMed=8408542; [<u>NCBI</u> , <u>ExPASy</u> , <u>EBI</u> , <u>Israel</u> ,

Japan]Campbell K., East A.K., Collins M.D.:

"Gene probes for identification of the botulinal neurotoxin gene and specific identification of neurotoxin types B, E, and F.";
J. Clin. Microbiol. 31:2255-2262(1993).

[4] SEQUENCE OF 1-44 AND 441-466.

STRAIN=657;

MEDLINE=89000987; PubMed=3139097; [NCBI, ExPASy, EBI, Israel, Japan]

Dasgupta B.R., Datta A.:

"Botulinum neurotoxin type B (strain 657): partial sequence and similarity with tetanus toxin.";
Biochimie 70:811-817(1988).

[5] SEQUENCE OF 1-16 AND 441-458.

STRAIN=OKRA;

MEDLINE=85197963; PubMed=3888113; [NCBI, ExPASy, EBI, Israel, Japan]

Schmidt J.J., Sathyamoorthy V., Dasgupta B.R.:

"Partial amino acid sequences of botulinum neurotoxins types B and E.";
Arch. Biochem. Biophys. 238:544-548(1985).

[6] IDENTIFICATION AS ZINC-PROTEASE.

MEDLINE=93054694; PubMed=1429690; [NCBI, ExPASy, EBI, Israel, Japan]

Schiavo G., Rossetto O., Santucci A., Dasgupta B.R., Montecucco C.:

"Botulinum neurotoxins are zinc proteins.";

J. Biol. Chem. 267:23479-23483(1992).

[7] IDENTIFICATION OF SUBSTRATE.

MEDLINE=93063293; PubMed=1331807; [NCBI, ExPASy, EBI, Israel, Japan]

Schiavo G., Benfenati F., Poulain B., Rossetto O., de Laureto P.P., Dasgupta B.R., Montecucco C.:

"Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin.";

Nature 359:832-835(1992).

Comments

FUNCTION: BOTULINUS TOXIN ACTS BY INHIBITING NEUROTRANSMITTER RELEASE. IT BINDS TO PERIPHERAL NEURONAL SYNAPSES, IS INTERNALIZED AND MOVES BY RETROGRADE TRANSPORT UP THE AXON INTO THE SPINAL CORD WHERE IT CAN MOVE BETWEEN POSTSYNAPTIC AND PRESYNAPTIC NEURONS. IT INHIBITS NEUROTRANSMITTER RELEASE BY ACTING AS A ZINC ENDOPEPTIDASE THAT CLEAVES THE 76-GLN-1-PHE-77 BOND OF SYNAPTOBREVIN-2.

CATALYTIC ACTIVITY: Limited hydrolysis of proteins of the neuroexocytosis apparatus, synaptobrevins, SNAP25 or syntaxin. No detected action on small molecule substrates.

COFACTOR: Binds 1 zinc ion per subunit (*By similarity*).

SUBUNIT: DISULFIDE-LINKED HETERODIMER OF A LIGHT CHAIN (L) AND A HEAVY CHAIN (H). THE LIGHT CHAIN HAS THE PHARMACOLOGICAL ACTIVITY, WHILE THE N-AND C-TERMINAL OF THE HEAVY CHAIN MEDIATE CHANNEL FORMATION AND TOXIN BINDING, RESPECTIVELY.

SUBCELLULAR LOCATION: Secreted.

MISCELLANEOUS: THERE ARE SEVEN ANTIGENICALLY DISTINCT FORMS OF BOTULINUM NEUROTOXIN: TYPES A, B, C1, D, E, F, AND G.

SIMILARITY: BELONGS TO PEPTIDASE FAMILY M27.

Copyright

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch)

Cross-references

EMBL	M81186; AAA23211.1; [<u>EMBL</u> / <u>GenBank</u> / <u>DDBJ</u>] -. [<u>CoDingSequence</u>] Z11934; CAA77991.1; [<u>EMBL</u> / <u>GenBank</u> / <u>DDBJ</u>] -. [<u>CoDingSequence</u>] X70817; CAA50148.1; [<u>EMBL</u> / <u>GenBank</u> / <u>DDBJ</u>] -. [<u>CoDingSequence</u>]
PIR	A48940; A48940.
PDB	1EPW; 01-NOV-00. [<u>ExPASy</u> / <u>RCSB</u>] 1F31; 01-NOV-00. [<u>ExPASy</u> / <u>RCSB</u>] 1F82; 16-AUG-00. [<u>ExPASy</u> / <u>RCSB</u>] 1F83; 16-AUG-00. [<u>ExPASy</u> / <u>RCSB</u>] 1FQH; 06-DEC-00. [<u>ExPASy</u> / <u>RCSB</u>] 1G9A; 13-NOV-02. [<u>ExPASy</u> / <u>RCSB</u>] 1G9B; 13-NOV-02. [<u>ExPASy</u> / <u>RCSB</u>] 1G9C; 13-NOV-02. [<u>ExPASy</u> / <u>RCSB</u>] 1G9D; 13-NOV-02. [<u>ExPASy</u> / <u>RCSB</u>] 1I1E; 21-NOV-01. [<u>ExPASy</u> / <u>RCSB</u>] <u>Detailed list of linked structures.</u>
MEROPS	<u>M27.002</u> ; -.
InterPro	<u>IPR000395</u> ; Bontoxilysin. <u>IPR006025</u> ; Zn_MTpeptdse. <u>Graphical view of domain structure.</u>
Pfam	<u>PF01742</u> ; Peptidase_M27; 1.
PRINTS	<u>PR00760</u> ; BONTOXILYSIN.
ProDom	<u>PD001963</u> ; Bontoxilysin; 1. <u>Domain structure</u> / <u>List of seq. sharing at least 1 domain</u>
PROSITE	<u>PS00142</u> ; ZINC_PROTEASE; 1.
HOBACGEN	[<u>Family</u> / <u>Alignment</u> / <u>Tree</u>]
BLOCKS	<u>P10844</u> .
ProtoNet	<u>P10844</u> .
ProtoMap	<u>P10844</u> .
PRESAGE	<u>P10844</u> .
DIP	<u>P10844</u> .

ModBase	P10844.
SWISS-2DPAGE	Get region on 2D PAGE.

Keywords

Neurotoxin; Transmembrane; Hydrolase; Metalloprotease; Zinc; 3D-structure.

Features

Key	From	To	Length	Description
INIT_MET	0	0		
CHAIN	1	440	440	BOTULINUM NEUROTOXIN B, LIGHT-CHAIN.
CHAIN	441	1290	850	BOTULINUM NEUROTOXIN B, HEAVY-CHAIN.
METAL	229	229		ZINC (CATALYTIC) (BY SIMILARITY).
ACT_SITE	230	230		BY SIMILARITY.
METAL	233	233		ZINC (CATALYTIC) (BY SIMILARITY).
DISULFID	436	445		INTERCHAIN (PROBABLE).
CONFLICT	29	29		T -> M (IN REF. 4).
CONFLICT	217	217		R -> G (IN REF. 2).
CONFLICT	224	224		A -> S (IN REF. 2).
CONFLICT	463	463		S -> R (IN REF. 4).

Sequence information

Length: 1290 AA [This is the length of the unprocessed precursor]	Molecular weight: 150670 Da [This is the MW of the unprocessed precursor]	CRC64: D21746E2C024DF43 [This is a checksum on the sequence]
--	--	---

10	20	30	40	50	60	
PVTINNFNYN	DPIDNNNIIM	MEPPFARGTG	RYYKAKFKITD	RIWIIPERYT	FGYKPEDFNK	
70	80	90	100	110	120	
SSGIFNRDVC	EYYDPDYLNNT	NDKKNIFLQT	MIKLFNRIKS	KPLGEKLLEM	IINGIPYLGD	
130	140	150	160	170	180	
RRVPLEEFNT	NIASVTVNKL	ISNPGEVERK	KGIFANLIIF	GPGPVLNENE	TIDIGIQNHF	

190	200	210	220	230	240
ASREGFGGIM	QMKFCPEYVS	VFNNVQENKG	ASIFNRRGYF	SDPALILMHE	LIHVLHGLYG
250	260	270	280	290	300
IKVDDLPIVP	NEKKFFMQST	DAIQAEELYT	FGGQDPSIIT	PSTDKSIYDK	VLQNFRGIVD
310	320	330	340	350	360
RLNKVLVCIS	DPNININIIYK	NKFKDKYKFKV	EDSEGKYSID	VESFDKLYKS	LMFGFTETNI
370	380	390	400	410	420
AENYKIKTRA	SYFSDSLPPV	KIKNLLDNEI	YTIEEGFNIS	DKDMEKEYRG	QNKAINKQAY
430	440	450	460	470	480
EEISKEHLAV	YKIQMCKSVK	APGICIDVDN	EDLFFIADKN	SFSDDLSSKNE	RIEYNTQSNSY
490	500	510	520	530	540
IENDFPINEL	ILDSDLISKI	ELPSENTESL	TDFNVDVPVY	EKQPAIKKIF	TDENTIFQYL
550	560	570	580	590	600
YSQTFPLDIR	DISLTSSFDD	ALLFSNKVYS	FFSMDYIKTA	NKVVEAGLFA	GWVKQIVNDF
610	620	630	640	650	660
VIEANKSNTM	DKIADISLIV	PYIGLALNVG	NETAKGNFEN	AFEIAGASIL	LEFIPELLIP
670	680	690	700	710	720
VVGAFILLESY	IDNKNKIIKT	IDNALTKRNE	KWSDMYGLIV	AQWLSTVNTQ	FYTIKEGMYK
730	740	750	760	770	780
ALNYQAQALE	EIIKYRYNIY	SEKEKSNINI	DFNDINSKLN	EGINQAIDNI	NNFINGCSV
790	800	810	820	830	840
YLMKKMIPLA	VEKLDFDNT	LKKNLLNYID	ENKLYLIGSA	EYEKSKVNVY	LKTIMPFDLS
850	860	870	880	890	900
IYTNDTILIE	MFNKYNSEIL	NNIILNLRYK	DNNLIDLDSGY	GAKVEVYDGV	ELNDKNQFKL
910	920	930	940	950	960
TSSANSKIRV	TQNQNIIFNS	VFLDFSVSFW	IRIPKYKNDG	IQNYIHNEYT	IINC MKNNSG
970	980	990	1000	1010	1020
WKISIRGNRI	IWTLDIDNGK	TKSVFFEYNI	REDISEYINR	WFFVTITNNL	NNAKIYINGK
1030	1040	1050	1060	1070	1080
LESNTDIKDI	REVIANGEII	FKLDGDDIDRT	QFIWMKYFSI	FNTELSQSNI	EERYKIQSYS
1090	1100	1110	1120	1130	1140
EYLKDFWGNP	LMYNKEYYMF	NAGNKNSYIK	LKKDSPVGEI	LTRSKYNQNS	KYINYRDLYI

1150	1160	1170	1180	1190	1200
GEKFIIRRKS	NSQSINDDIV	RKEDYIYLDF	FNLNQEWRVY	TYKYFKKEEE	KLFLAPISDS
1210	1220	1230	1240	1250	1260
DEFYNTIQIK	EYDEQPTYSC	QLLFKKDEES	TDEIGLIGH	RFYESGIVFE	EYKDYFCISK
1270	1280	1290			
WYLKEVKRKP	YNLKLGCNWQ	FIPKDEGWTE			

P10844 in
FASTA format

[View entry in original Swiss-Prot format](#)

[View entry in raw text format \(no links\)](#)

[Report form for errors/updates in this Swiss-Prot entry](#)

BLAST [BLAST submission on](#)
[ExPASy/SIB](#)
 or at [NCBI \(USA\)](#)



Sequence analysis tools: [ProtParam](#),
[ProtScale](#), [Compute pI/Mw](#),
[PeptideMass](#), [PeptideCutter](#),
[Dotlet](#) (Java)



[ScanProsite](#),
[MotifScan](#)



Search the [SWISS-MODEL](#)
[Repository](#)

[ExPASy Home page](#) [Site Map](#) [Search ExPASy](#) [Contact us](#) [Swiss-Prot](#)

[Hosted by NCSC US](#) [Mirror sites:](#) [Canada](#) [China](#) [Korea](#) [Switzerland](#) [Taiwan](#)

ExPASy Home page	Site Map	Search ExPASy	Contact us	Swiss-Prot
Search	Swiss-Prot/TrEMBL	<input type="button" value="▼ for"/>	<input type="text"/>	<input type="button" value="Go"/> <input type="button" value="Clear"/>

Swiss-Prot: BXB_CLOBO

The section of the sequence BXB_CLOBO (P10844) you have selected corresponds to:

SEQUENCE 1 44

In one-letter code:

1	11	21	31	41	51	
1 PVTINNFNYN	DPIDNNNIIM	MEPPFARGTG	RYYKAFKITD	RIWIIPERYT	FGYKPEDFNK	60
61 SSGIFNRDVC	EYYDPDYLNT	NDKKNIFLQT	MIKLFNRIKS	KPLGEKLLEM	IINGIPYLGD	120
121 RRVPLEEFNT	NIASVTVNKL	ISNPGEVERK	KGIFANLIIF	GPGPVLNENE	TIDIGIQNHF	180
181 ASREGFGGIM	QMFKCPEYVS	VFNNVQENKG	ASIFNRRGYF	SDPALILMHE	LIHVLHGLYG	240
241 IKVDDLPPIVP	NEKKFFMQST	DAIQAEELYT	FGGQDPSIIT	PSTDKSIYDK	VLQNFRGIVD	300
301 RLNKVLVCIS	DPNININIYK	NKFKDKYKFV	EDSEGKYSID	VESFDKLYKS	LMFGFTETNI	360
361 AENYKIKTRA	SYFDSLPPV	KIKNLLDNEI	YTIEEGFNIS	DKDMEKEYRG	QNKAINQAY	420
421 EEISKEHLAV	YKIQMCKSVK	APGICIDVDN	EDLFFIADKN	SFSDDLSKNE	RIEYNTQSNSY	480
481 IENDFPINEL	ILDSDLISKI	ELPSENTESL	TDFNVDVPVY	EKQPAIKKIF	TDENTIFQYL	540
541 YSQTFPPLDIR	DISLTSSFDD	ALLFSNKVYS	FFSMDYIKTA	NKVVEAGLFA	GWVKQIVNDF	600
601 VIEANKSNTM	DKIADISLIV	PYIGLALNVG	NETAKGNFEN	AFEIAGASIL	LEFIPELLIP	660
661 VVGAFLLESY	IDNKNKIIKT	IDNALTKRNE	KWSDMYGLIV	AQWLSTVNTQ	FYTIKEGMYK	720
721 ALNYQAQALE	EIIKYRYNIY	SEKEKSNINI	DFNDINSKLN	EGINQAIDNI	NNFINGCSV	780
781 YLMKKMIPLA	VEKLLDFDNT	LKKNLLNYID	ENKLYLIGSA	EYEKSKVNKY	LKTIMPFDLS	840
841 IYTNDTILIE	MFNKYNSEIL	NNIILNLRYK	DNNLIDLSGY	GAKVEVYDGV	ELNDKNQFKL	900
901 TSSANSKIRV	TQNQNIIFNS	VFLDFSVFW	IRIPKYKNDG	IQNYIHNEYT	IINCMKNNSG	960
961 WKISIRGNRI	IWTLDINGK	TKSVFFEYNI	REDISEYINR	WFFVTITNNL	NNAKIYINGK	1020
1021 LESNTDIKDI	REVIANGEII	FKLDGDDIDRT	QFIWMKYFSI	FNTELSQSNSI	EERYKIQSYS	1080
1081 EYLKDFWGPNP	LMYNKEYYMF	NAGNKNSYIK	LKKDSPVGEI	LTRSKYNQNS	KYINYRDLYI	1140
1141 GEKFIIRRKS	NSQSINDDIV	RKEDYIYLDF	FNLNQEWRVY	TYKYFKKEEE	KLFLAPISDS	1200
1201 DEFYNTIQIK	EYDEQPTYSC	QLLFFKKDEES	TDEIGLIGH	RFYESGIVFE	EYKDYFCISK	1260
1261 WYLKEVKRKP	YNLKLGCNWQ	FIPKDEGWTE				

In three-letter code:

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1 Pro	Val	Thr	Ile	Asn	Asn	Phe	Asn	Tyr	Asn	Asp	Pro	Ile	Asp	Asn	15
16 Asn	Asn	Ile	Ile	Met	Met	Glu	Pro	Pro	Phe	Ala	Arg	Gly	Thr	Gly	30
31 Arg	Tyr	Tyr	Lys	Ala	Phe	Lys	Ile	Thr	Asp	Arg	Ile	Trp	Ile	Ile	45
46 Pro	Glu	Arg	Tyr	Thr	Phe	Gly	Tyr	Lys	Pro	Glu	Asp	Phe	Asn	Lys	60
61 Ser	Ser	Gly	Ile	Phe	Asn	Arg	Asp	Val	Cys	Glu	Tyr	Tyr	Asp	Pro	75
76 Asp	Tyr	Leu	Asn	Thr	Asn	Asp	Lys	Lys	Asn	Ile	Phe	Leu	Gln	Thr	90
91 Met	Ile	Lys	Leu	Phe	Asn	Ile	Lys	Ser	Lys	Pro	Leu	Gly	Glu		105

106	Lys	Leu	Leu	Glu	Met	Ile	Ile	Asn	Gly	Ile	Pro	Tyr	Leu	Gly	Asp	120
121	Arg	Arg	Val	Pro	Leu	Glu	Glu	Phe	Asn	Thr	Asn	Ile	Ala	Ser	Val	135
136	Thr	Val	Asn	Lys	Leu	Ile	Ser	Asn	Pro	Gly	Glu	Val	Glu	Arg	Lys	150
151	Lys	Gly	Ile	Phe	Ala	Asn	Leu	Ile	Ile	Phe	Gly	Pro	Gly	Pro	Val	165
166	Leu	Asn	Glu	Asn	Glu	Thr	Ile	Asp	Ile	Gly	Ile	Gln	Asn	His	Phe	180
181	Ala	Ser	Arg	Glu	Gly	Phe	Gly	Gly	Ile	Met	Gln	Met	Lys	Phe	Cys	195
196	Pro	Glu	Tyr	Val	Ser	Val	Phe	Asn	Asn	Val	Gln	Glu	Asn	Lys	Gly	210
211	Ala	Ser	Ile	Phe	Asn	Arg	Arg	Gly	Tyr	Phe	Ser	Asp	Pro	Ala	Leu	225
226	Ile	Leu	Met	His	Glu	Leu	Ile	His	Val	Leu	His	Gly	Leu	Tyr	Gly	240
241	Ile	Lys	Val	Asp	Asp	Leu	Pro	Ile	Val	Pro	Asn	Glu	Lys	Lys	Phe	255
256	Phe	Met	Gln	Ser	Thr	Asp	Ala	Ile	Gln	Ala	Glu	Glu	Leu	Tyr	Thr	270
271	Phe	Gly	Gly	Gln	Asp	Pro	Ser	Ile	Ile	Thr	Pro	Ser	Thr	Asp	Lys	285
286	Ser	Ile	Tyr	Asp	Lys	Val	Leu	Gln	Asn	Phe	Arg	Gly	Ile	Val	Asp	300
301	Arg	Leu	Asn	Lys	Val	Leu	Val	Cys	Ile	Ser	Asp	Pro	Asn	Ile	Asn	315
316	Ile	Asn	Ile	Tyr	Lys	Asn	Lys	Phe	Lys	Asp	Lys	Tyr	Lys	Phe	Val	330
331	Glu	Asp	Ser	Glu	Gly	Lys	Tyr	Ser	Ile	Asp	Val	Glu	Ser	Phe	Asp	345
346	Lys	Leu	Tyr	Lys	Ser	Leu	Met	Phe	Gly	Phe	Thr	Glu	Thr	Asn	Ile	360
361	Ala	Glu	Asn	Tyr	Lys	Ile	Lys	Thr	Arg	Ala	Ser	Tyr	Phe	Ser	Asp	375
376	Ser	Leu	Pro	Pro	Val	Lys	Ile	Lys	Asn	Leu	Leu	Asp	Asn	Glu	Ile	390
391	Tyr	Thr	Ile	Glu	Gly	Phe	Asn	Ile	Ser	Asp	Lys	Asp	Met	Glu	405	
406	Lys	Glu	Tyr	Arg	Gly	Gln	Asn	Lys	Ala	Ile	Asn	Lys	Gln	Ala	Tyr	420
421	Glu	Glu	Ile	Ser	Lys	Glu	His	Leu	Ala	Val	Tyr	Lys	Ile	Gln	Met	435
436	Cys	Lys	Ser	Val	Lys	Ala	Pro	Gly	Ile	Cys	Ile	Asp	Val	Asp	Asn	450
451	Glu	Asp	Leu	Phe	Phe	Ile	Ala	Asp	Lys	Asn	Ser	Phe	Ser	Asp	Asp	465
466	Leu	Ser	Lys	Asn	Glu	Arg	Ile	Glu	Tyr	Asn	Thr	Gln	Ser	Asn	Tyr	480
481	Ile	Glu	Asn	Asp	Phe	Pro	Ile	Asn	Glu	Leu	Ile	Leu	Asp	Thr	Asp	495
496	Leu	Ile	Ser	Lys	Ile	Glu	Leu	Pro	Ser	Glu	Asn	Thr	Glu	Ser	Leu	510
511	Thr	Asp	Phe	Asn	Val	Asp	Val	Pro	Val	Tyr	Glu	Lys	Gln	Pro	Ala	525
526	Ile	Lys	Lys	Ile	Phe	Thr	Asp	Glu	Asn	Thr	Ile	Phe	Gln	Tyr	Leu	540
541	Tyr	Ser	Gln	Thr	Phe	Pro	Leu	Asp	Ile	Arg	Asp	Ile	Ser	Leu	Thr	555
556	Ser	Ser	Phe	Asp	Asp	Ala	Leu	Leu	Phe	Ser	Asn	Lys	Val	Tyr	Ser	570
571	Phe	Phe	Ser	Met	Asp	Tyr	Ile	Lys	Thr	Ala	Asn	Lys	Val	Val	Glu	585
586	Ala	Gly	Leu	Phe	Ala	Gly	Trp	Val	Lys	Gln	Ile	Val	Asn	Asp	Phe	600
601	Val	Ile	Glu	Ala	Asn	Lys	Ser	Asn	Thr	Met	Asp	Lys	Ile	Ala	Asp	615
616	Ile	Ser	Leu	Ile	Val	Pro	Tyr	Ile	Gly	Leu	Ala	Leu	Asn	Val	Gly	630
631	Asn	Glu	Thr	Ala	Lys	Gly	Asn	Phe	Glu	Asn	Ala	Phe	Glu	Ile	Ala	645
646	Gly	Ala	Ser	Ile	Leu	Glu	Phe	Ile	Pro	Glu	Leu	Leu	Ile	Pro		660
661	Val	Val	Gly	Ala	Phe	Leu	Leu	Glu	Ser	Tyr	Ile	Asp	Asn	Lys	Asn	675
676	Lys	Ile	Ile	Lys	Thr	Ile	Asp	Asn	Ala	Leu	Thr	Lys	Arg	Asn	Glu	690
691	Lys	Trp	Ser	Asp	Met	Tyr	Gly	Leu	Ile	Val	Ala	Gln	Trp	Leu	Ser	705
706	Thr	Val	Asn	Thr	Gln	Phe	Tyr	Thr	Ile	Lys	Glu	Gly	Met	Tyr	Lys	720
721	Ala	Leu	Asn	Tyr	Gln	Ala	Gln	Ala	Leu	Glu	Glu	Ile	Ile	Lys	Tyr	735
736	Arg	Tyr	Asn	Ile	Tyr	Ser	Glu	Lys	Glu	Lys	Ser	Asn	Ile	Asn	Ile	750
751	Asp	Phe	Asn	Asp	Ile	Asn	Ser	Lys	Leu	Asn	Glu	Gly	Ile	Asn	Gln	765
766	Ala	Ile	Asp	Asn	Ile	Asn	Asn	Phe	Ile	Asn	Gly	Cys	Ser	Val	Ser	780
781	Tyr	Leu	Met	Lys	Lys	Met	Ile	Pro	Leu	Ala	Val	Glu	Lys	Leu	Leu	795
796	Asp	Phe	Asp	Asn	Thr	Leu	Lys	Lys	Asn	Leu	Leu	Asn	Tyr	Ile	Asp	810
811	Glu	Asn	Lys	Leu	Tyr	Leu	Ile	Gly	Ser	Ala	Glu	Tyr	Glu	Lys	Ser	825
826	Lys	Val	Asn	Lys	Tyr	Leu	Lys	Thr	Ile	Met	Pro	Phe	Asp	Leu	Ser	840
841	Ile	Tyr	Thr	Asn	Asp	Thr	Ile	Leu	Ile	Glu	Met	Phe	Asn	Lys	Tyr	855
856	Asn	Ser	Glu	Ile	Leu	Asn	Asn	Ile	Ile	Leu	Asn	Leu	Arg	Tyr	Lys	870
871	Asp	Asn	Asn	Leu	Ile	Asp	Leu	Ser	Gly	Tyr	Gly	Ala	Lys	Val	Glu	885
886	Val	Tyr	Asp	Gly	Val	Glu	Leu	Asn	Asp	Lys	Asn	Gln	Phe	Lys	Leu	900
901	Thr	Ser	Ser	Ala	Asn	Ser	Lys	Ile	Arg	Val	Thr	Gln	Asn	Gln	Asn	915
916	Ile	Ile	Phe	Asn	Ser	Val	Phe	Leu	Asp	Phe	Ser	Val	Ser	Phe	Trp	930
931	Ile	Arg	Ile	Pro	Lys	Tyr	Lys	Asn	Asp	Gly	Ile	Gln	Asn	Tyr	Ile	945
946	His	Asn	Glu	Tyr	Thr	Ile	Ile	Asn	Cys	Met	Lys	Asn	Asn	Ser	Gly	960
961	Trp	Lys	Ile	Ser	Ile	Arg	Gly	Asn	Arg	Ile	Ile	Trp	Thr	Leu	Ile	975
976	Asp	Ile	Asn	Gly	Lys	Thr	Lys	Ser	Val	Phe	Phe	Glu	Tyr	Asn	Ile	990
991	Arg	Glu	Asp	Ile	Ser	Glu	Tyr	Ile	Asn	Arg	Trp	Phe	Phe	Val	Thr	1005
1006	Ile	Thr	Asn	Asn	Leu	Asn	Asn	Ala	Lys	Ile	Tyr	Ile	Asn	Gly	Lys	1020
1021	Leu	Glu	Ser	Asn	Thr	Asp	Ile	Lys	Asp	Ile	Arg	Glu	Val	Ile	Ala	1035
1036	Asn	Gly	Glu	Ile	Ile	Phe	Lys	Leu	Asp	Gly	Asp	Ile	Asp	Arg	Thr	1050
1051	Gln	Phe	Ile	Trp	Met	Lys	Tyr	Phe	Ser	Ile	Phe	Asn	Thr	Glu	Leu	1065
1066	Ser	Gln	Ser	Asn	Ile	Glu	Glu	Arg	Tyr	Lys	Ile	Gln	Ser	Tyr	Ser	1080

1081	Glu	Tyr	Leu	Lys	Asp	Phe	Trp	Gly	Asn	Pro	Leu	Met	Tyr	Asn	Lys	1095
1096	Glu	Tyr	Tyr	Met	Phe	Asn	Ala	Gly	Asn	Lys	Asn	Ser	Tyr	Ile	Lys	1110
1111	Leu	Lys	Lys	Asp	Ser	Pro	Val	Gly	Glu	Ile	Leu	Thr	Arg	Ser	Lys	1125
1126	Tyr	Asn	Gln	Asn	Ser	Lys	Tyr	Ile	Asn	Tyr	Arg	Asp	Leu	Tyr	Ile	1140
1141	Gly	Glu	Lys	Phe	Ile	Ile	Arg	Arg	Lys	Ser	Asn	Ser	Gln	Ser	Ile	1155
1156	Asn	Asp	Asp	Ile	Val	Arg	Lys	Glu	Asp	Tyr	Ile	Tyr	Leu	Asp	Phe	1170
1171	Phe	Asn	Leu	Asn	Gln	Glu	Trp	Arg	Val	Tyr	Thr	Tyr	Lys	Tyr	Phe	1185
1186	Lys	Lys	Glu	Glu	Lys	Leu	Phe	Leu	Ala	Pro	Ile	Ser	Asp	Ser	1200	
1201	Asp	Glu	Phe	Tyr	Asn	Thr	Ile	Gln	Ile	Lys	Glu	Tyr	Asp	Glu	Gln	1215
1216	Pro	Thr	Tyr	Ser	Cys	Gln	Leu	Leu	Phe	Lys	Lys	Asp	Glu	Glu	Ser	1230
1231	Thr	Asp	Glu	Ile	Gly	Leu	Ile	Gly	Ile	His	Arg	Phe	Tyr	Glu	Ser	1245
1246	Gly	Ile	Val	Phe	Glu	Glu	Tyr	Lys	Asp	Tyr	Phe	Cys	Ile	Ser	Lys	1260
1261	Trp	Tyr	Leu	Lys	Glu	Val	Lys	Arg	Lys	Pro	Tyr	Asn	Leu	Lys	Leu	1275
1276	Gly	Cys	Asn	Trp	Gln	Phe	Ile	Pro	Lys	Asp	Glu	Gly	Trp	Thr	Glu	

Direct similarity search submission of this subsequence to

BLAST

[BLAST submission on](#)
[ExPASy/SIB](#)
 or at [NCBI \(USA\)](#)



Sequence analysis tools:
[ProtParam](#), [ProtScale](#), [Compute pI/Mw](#), [PeptideMass](#),
[PeptideCutter](#), [Dotlet \(Java\)](#)



[ScanProsite](#)



[Direct Submission to](#)
[SWISS-MODEL](#)

ID CBBONT standard; DNA; PRO; 4041 BP.
 XX
 AC M81186;
 XX
 SV M81186.1
 XX
 DT 28-MAY-1992 (Rel. 32, Created)
 DT 04-MAR-2000 (Rel. 63, Last updated, Version 4)
 XX
 DE Clostridium botulinum neurotoxin type B (botB) gene, complete cds.
 XX
 KW botB gene; neurotoxin type B.
 XX
 OS Clostridium botulinum
 OC Bacteria; Firmicutes; Clostridia; Clostridiales; Clostridiaceae;
 OC Clostridium.
 XX
 RN [1]
 RP 1-4041
 RA Whelan S.M., Elmore M.J., Bodsworth N.J., Brehm J.K., Atkinson T.,
 RA Minton N.P.;
 RT "Complete nucleotide sequence of the Clostridium botulinum gene encoding
 RT the type B neurotoxin";
 RL Unpublished.
 XX
 DR GOA; P10844; P10844.
 DR SWISS-PROT; P10844; BXB_CLOBO.
 XX
 FH Key Location/Qualifiers
 FH
 FT source 1..4041
 FT /db_xref="taxon:1491"
 FT /organism="Clostridium botulinum"
 FT CDS 57..3932
 FT /codon_start=1
 FT /db_xref="GOA:P10844"
 FT /db_xref="SWISS-PROT:P10844"
 FT /transl_table=11
 FT /function="vertebrate neurotoxin"
 FT /product="neurotoxin type B"
 FT /gene="botB"
 FT /protein_id="AAA23211.1"
 FT /translation="MPVTINNFNYNDPIDNNNIIMMEPPFARGTGRYYKAFKITDRIWI
 IPERYTFGYKPEDFNKSSGIFNRDVCEYYDPDYLNTNDKKNIFLQTMIKLFNRRIKS
 GEKLLLEMIINGIPYLGDRRVPLEEFNTNIASTVNKLISNPGEVERKKGIFANLIIFGP
 GPVLNESETIDIGIQNHFASREGFGGIMQMFKCPEYVSFNNVQENKGASIFNRGGYFS
 DPALILMHeliHVHLGLYGIKVDDLPIVPNEKKFMQSTD
 DAQAEELYTFGGQDPSIIT
 PSTDKSIYDKVLQNF
 RGVIVDRLNKLV
 CISDPNININ
 IYKNKF
 KDKYKF
 VEDSEGKYSI
 DVESFDKLYKSLMFGFT
 ETNIAENYKIKTRASYFS
 DLSLPPVK
 IKNLLDNEIY
 TIEEGFN
 ISDKDMEKEYRGQ
 QNAINKQAYEE
 ISKEHLAVY
 KIQMC
 KS
 VAPGIC
 IDVD
 NEDLF
 FIA
 DKNSFS
 DDL
 SKNERIE
 YNTQS
 NYIEND
 FPI
 NEL
 IL
 DTL
 I
 SKIELP
 S
 ENTES
 LTDF
 NVD
 VPVYE
 KQPA
 I
 KK
 IF
 DENT
 IF
 QYLYS
 QTFPL
 DIR
 DIS
 LTSS
 FDD
 ALL
 FS
 NKV
 YSF
 SMD
 YIKTANK
 VVEAGL
 FAGWVK
 QIVNDFV
 IE
 BANKS
 NTMD
 KIA
 DISL
 IVPY
 IGL
 ALNV
 GNETA
 KGNF
 ENAF
 EIA
 GAS
 S
 ILL
 F
 IPE
 L
 PV
 V
 GA
 F
 L
 E
 SY
 IDN
 KN
 K
 I
 K
 T
 D
 N
 A
 L
 T
 K
 R
 N
 E
 K
 W
 S
 D
 M
 Y
 G
 L
 I
 V
 A
 Q
 W
 L
 S
 T
 V
 N
 T
 Q
 F
 Y
 T
 I
 K
 E
 G
 M
 Y
 K
 A
 L
 N
 Y
 Q
 A
 L
 E
 E
 I
 I
 K
 Y
 R
 N
 I
 Y
 S
 E
 K
 E
 K
 S
 N
 I
 N
 D
 F
 I
 N
 D
 S
 K
 I
 L
 N
 E
 G
 I
 N
 Q
 A
 I
 D
 N
 I
 N
 N
 F
 I
 N
 G
 C
 S
 V
 Y
 L
 M
 K
 M
 I
 P
 L
 A
 V
 E
 K
 L
 L
 D
 F
 N
 T
 L
 K
 K
 N
 L
 N
 Y
 I
 D
 E
 N
 K
 L
 Y
 L
 I
 G
 S
 A
 E
 Y
 K
 S
 V
 N
 K
 Y
 L
 K
 T
 I
 M
 P
 F
 D
 L
 S
 I
 Y
 T
 N
 D
 T
 I
 L
 I
 E
 M
 F
 N
 K
 Y
 N
 S
 E
 I
 L
 N
 N
 I
 I
 L
 N
 L
 R
 Y
 K
 D
 N
 N
 L
 I
 D
 L
 S
 G
 Y
 G
 A
 K
 V
 E
 V
 D
 G
 V
 E
 L
 N
 D
 K
 Q
 F
 K
 L
 T
 S
 A
 N
 S
 K
 I
 R
 V
 T
 Q
 N
 Q
 N
 I
 I
 F
 N
 S
 V
 F
 L
 D
 F
 S
 V
 F
 S
 W
 R
 I
 P
 K
 Y
 K
 N
 D
 G
 I
 Q
 N
 Y
 I
 H
 N
 E
 T
 I
 I
 N
 C
 M
 K
 N
 S
 G
 W
 K
 I
 S
 I
 R
 G
 N
 R
 I
 I
 W
 T
 L
 I
 D
 I
 N
 G
 K
 T
 S
 V
 F
 E
 Y
 N
 I
 R
 E
 D
 I
 S
 E
 Y
 I
 N
 R
 W
 F
 F
 V
 T
 I
 T
 N
 N
 L
 N
 A
 K
 I
 Y
 I
 N
 G
 K
 L
 E
 S
 N
 T
 D
 I
 K
 D
 I
 R
 E
 V
 I
 A
 N
 G
 E
 I
 I
 F
 K
 L
 D
 G
 D
 I
 D
 R
 T
 Q
 F
 I
 W
 M
 K
 Y
 F
 S
 I
 F
 N
 T
 E
 L
 S
 Q
 N
 I
 E
 E
 R
 Y
 K
 I
 Q
 S
 Y
 S
 E
 Y
 L
 K
 D
 F
 W
 G
 N
 P
 L
 M
 Y
 N
 K
 E
 Y
 Y
 M
 F
 N
 A
 G
 N
 K
 N
 S
 I
 N
 D
 D
 I
 V
 R
 K
 D
 E
 E
 S
 T
 D
 E
 I
 G
 L
 I
 G
 H
 R
 F
 Y
 E
 S
 G
 I
 V
 F
 E
 E
 Y
 K
 D
 Y
 F
 C
 I
 S
 K
 W
 Y
 L
 K
 E
 V
 K
 R
 K
 P
 Y
 N
 L
 K
 G
 C
 N
 W
 Q
 F
 I
 P
 K
 D
 E
 G
 W
 T
 E

updated
5/03/03
KD

XX		
SQ	Sequence 4041 BP; 1679 A; 383 C; 645 G; 1334 T; 0 other;	
	tgcgcattta tggcattaa aaggatata aactaaaaat aaggaggaga atatttatgc	60
	cagttacaat aaataatttt aattataatg atccattga taataataat attattatga	120
	tggagctcc atttgcaga ggtacgggaa gatattataa agctttaaa atcacagatc	180
	gtatggat aataccgaa agatatactt ttggatataa acctgaggat tttaataaaa	240
	gttccgtat tttaataga gatgttgat aatattatga tccagattac ttaaatacta	300
	atgataaaaaa gaatatattt ttacaaacaa tgatcaagtt attaataga atcaaataaa	360
	aaccattggg tggaaagtta ttagagatga ttataatgg tatacctt atggagata	420
	gacgtttcc actcgaagag tttaacacaa acattgctag tgtaactgtt aataatataa	480
	ttagtaatcc aggagaagtg gagcgaaaaa aaggatattt cgcaaattta ataataatgg	540
	gacctggggc agtttaaat gaaaatgaga ctatagatat aggtatacaa aatcattttg	600
	catcaagggg aggcttcggg ggtataatgc aaatgaagtt ttgcccagaa tatgtaaagcg	660
	tatataataa tggtaagaa aacaaaggcg caagtatatt taatagacgt ggatatttt	720
	cagatccagc cttgatatta atgcataac ttatacatgt ttacatgga ttatatggca	780
	ttaaagtaga tgatttacca attgtaccaa atgaaaaaaa atttttatg caatctacag	840
	atgctataca ggcagaagaa ctatatacat ttggaggaca agatcccagc atataactc	900
	cttctacgga taaaagtatc tatgataaaag ttttgc当地 ttttagaggg atagttgata	960
	gacttaacaa ggtttagtt tgcatatcatc atccatacat taatattaat atatataaaa	1020
	ataaaatataa agataaaatataa aattcgtt aagattctga gggaaaatata agtataatgt	1080
	tagaaagttt tgataaaatta tataaaagct taatgtttgg ttttacagaa actaatatag	1140
	cagaaaatataa taaaataaaa actagagctt cttatTTT tagttccctta ccaccagtaa	1200
	aaataaaaaaaa ttatttagat aatgaaatct atactataga ggaagggttt aatataatctg	1260
	ataaaagatataa gggaaaagaa tatagaggtc agaataaagc tataaataaaa caagcttatg	1320
	aagaaatttag caaggagcat ttggctgtat ataagataca aatgtgtaaa agtgtttaag	1380
	ctccaggaat atgtattgtat gttgataatg aagatttgc ttttatactg gataaaaata	1440
	gtttttcaga tgatttatct aaaaacgaaa gaatagaata taatacacag agtaattata	1500
	tagaaaatataa cttccctata aatgaattaa ttttagatac tgatttaata agtaaaatata	1560
	aattaccaag tggaaatataca gaatcacca ctgattttaa ttttagatgtt ccagtatatg	1620
	aaaaacacaacc cgctataaaa aaaatttttta cagatggaaa taccatctt caatatttt	1680
	actctcagac atttcctcta gatataagag atataagttt aacatcttca tttgatgtat	1740
	cattattatt ttctaaacaaa gtttatttcat tttttctat ggattatatt aaaactgcta	1800
	ataaaagtgtt agaagcagga ttatttgcag gttgggtgaa acagatagta aatgattttg	1860
	taatcgaagc taataaaagc aatactatgg ataaaattgc agatataatct ctaattgttc	1920
	cttatataagg attagctta aatgttagaa atgaaacagc taaaggaaat ttgaaaatgt	1980
	cttttgagat tgccggagcc agtattctac tagaattttt accagaactt ttaatacctg	2040
	tagttggagc ctttttattta gaatcatata ttgacaataa aaataaaaattt attaaaacaa	2100
	tagataatgc tttaactaaa agaaatgaaa aatggagtga tatgtacgaa ttaatagtag	2160
	cgcatacgct ctcaacagtt aatactcaat ttatacaat aaaagggaa atgtataagg	2220
	ctttaaatataa tcaagcacaa gcattggaaat aaataataaa atacagatata aatataatatt	2280
	ctgaaaaaaga aaagtcaaat attaacatcg attttatgt tataaattct aaacttaatg	2340
	agggtattaa ccaagctata gataatataa ataattttttaa aatggatgt tctgtatcat	2400
	atttaatgaa aaaaatgatt ccattagctg tagaaaaattt actagactt gataatactc	2460
	tcaaaaaaaaaa ttgtttaat tatataatgt aaaaataattt atatttgatt ggaagtgcag	2520
	aatatgaaaaa atcaaaaatgt aataaaatact tgaaaaccat tatgccgtt gatcttcaa	2580
	tatataccaa tgatacaata ctaatgaaaaa tggttaataa atataatagc gaaattttaa	2640
	ataatattat cttaaattta agatataaagg ataataattt aatagattta tcaggatatg	2700
	ggccaaaggtt agaggatata gatggagtc agcttaatgtg taaaatcaa tttaaattaa	2760
	ctagttcagc aaatagtaag attagatgtc ctcaaaatca gaatatcata tttaatagtg	2820
	tgttccttgc tttagcgtt agcttttgc taagaatacc taaatataag aatgtatggta	2880
	tacaaaattta tattcataat gaatatacaa taattatgt tatggaaaat aattcgggct	2940
	gaaaaatatac tattagggtt aataggataa tatggacttt aattgtatata aatggaaaaaa	3000
	ccaaatcggtt attttttgaa tataacataa gagaagatatac atcagatata aataatagat	3060
	gttttttgtt aactattact aataatttgc ataaacgctaa aatttattatt aatggtaagc	3120
	tagaatcaaa tacagatattt aatggatataa gagaagttt tgctaatgtt gaaataatata	3180
	ttaaatttgc ttgttgcata gatagaacac aattttatgt gatgaatata ttcatgtatt	3240
	ttaatacggaa attaagtcaat ctaaaatattt aagaaagata taaaattcaaa tcataatagcg	3300
	aatattttaa agattttgg gggaaatccctt taatgtacaa taaagaatata tatgttttta	3360
	atgcggggaa taaaattca tatattaaac taaagaaaga ttccacgtt ggtggaaat	3420
	taacacgttca caaatataat caaaattcta aatataaaaa ttatagatgtt ttatataattt	3480
	gagaaaaattt tattataaga agaaagtcaat attctcaatc tataaattgtat gatatagtt	3540
	gaaaagaaga ttatataat ctagattttt ttaattttaa tcaagagtgg agagtatata	3600
	cctataaataa tttaagaaa gaggaagaaa aattgtttttt agctccata agtgattctg	3660
	atgagtttca caatactata caaataaaag aatgtatgtca acagccaaaca tatgtttgtc	3720
	agttgcatttt taaaaatgtt gaaagaaatgtt ctgtatgatgtt aggattgtt ggtatttcatc	3780

gtttctacga atcttggatt gtatTTGAAG agtataAAAGA ttatTTTGT ataagtaaaAT	3840
ggTACTTAAAG agaggtaaaa aggaaaccat ataatttaaa attgggatgt aattggcagt	3900
ttattcctaa agatgaaggg tggactgaat aatataacta tatgctcagc aaaccttattt	3960
tatataagaa aagttaagt ttataaaatc ttaagtttaa ggatgttagct aaatTTTgaa	4020
tattagataa actacatgtt T	4041

//